VERIFICATION ASYMMETRY IN OPTIMAL TRANSPORT

Anonymous authors
Paper under double-blind review

ABSTRACT

Optimal transport (OT) solvers often return a coupling that is subsequently used as a correspondence, alignment, or training target. We study a question distinct from computing that coupling: given a candidate and a declared proof language, how much evidence is needed to certify near-optimality or refute it? We introduce a candidate-specific, proof-language-relative verification profile with two sides: the interaction width of a positive certificate and the size of a replayable negative witness. These requirements can scale differently. Under coordinate-block proofs, a d-dimensional parity family requires full width d to certify any optimal candidate to tolerance below 4, while every full-support suboptimal candidate has a four-edge refutation. The positive obstruction persists under Wasserstein perturbations; on the negative side, some four-cycle has unit improvement margin at least half of the global suboptimality gap. We also show how local Kantorovich duals compose into global lower bounds, when certificate gaps control plan error, and when incomplete observations force abstention. VerifyOT returns CERTIFIED, REFUTED, or UNKNOWN. Theorem-aligned experiments recover the predicted width regimes and four-cycle bound. On Sinkhorn/Greenkhorn candidates, positive closure appears near the exact-OT regime, while every repaired full-support output in our audit admits an exact four-cycle refutation under the common unregularized cost.

1 INTRODUCTION

Optimal transport has two outputs that are easy to conflate: an optimal value and an optimal coupling. The value is used as a discrepancy or loss; the coupling is used as an alignment, pseudo-label assignment, interpolation rule, or minibatch pairing. In the second use case, a plausible objective value is not enough. A downstream user needs evidence that the returned coupling is near-optimal under the stated cost, or a concrete mass reassignment showing that it is not.

Classical OT theory already provides complete optimality criteria. A feasible plan and a feasible Kantorovich dual pair yield a primal–dual gap, and minimum-cost-flow theory characterizes optimality by the absence of negative residual cycles (Kantorovich, 1942; Ahuja et al., 1993; Villani, 2009). Our question is different:

Given a candidate OT coupling and a declared proof language, how much local evidence is needed to certify its near-optimality or refute it, and can these two evidence requirements scale differently?

This question is relevant because modern ML pipelines routinely use approximate transport couplings rather than only OT values. Entropic and greedy scaling methods trade exactness for speed (Cuturi, 2013; Altschuler et al., 2017; Genevay et al., 2019); minibatch methods solve collections of smaller transport problems (Fatras et al., 2020; Nguyen et al., 2022); low-rank methods restrict the coupling class (Forrow et al., 2019; Scetbon et al., 2021; Scetbon & Cuturi, 2022); and neural or amortized methods learn transport maps or potentials (Makkuva et al., 2020; Asadulaev et al., 2024; Truong & Nguyen, 2026). We therefore treat verification as a separate, candidate-specific problem and make the admissible *proof language* explicit. Quadratic Wasserstein OT is especially useful here because orthogonal cost decomposition supports positive certificates, Kantorovich duality makes them rigorous, and the squared-Euclidean cross-difference identity turns negative evidence into local geometry. Related work is discussed in Section 8.

Three contributions. (1) A bidirectional, proof-language-relative notion of verification complexity. We pair positive verification width with negative refutation size and use this profile to

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

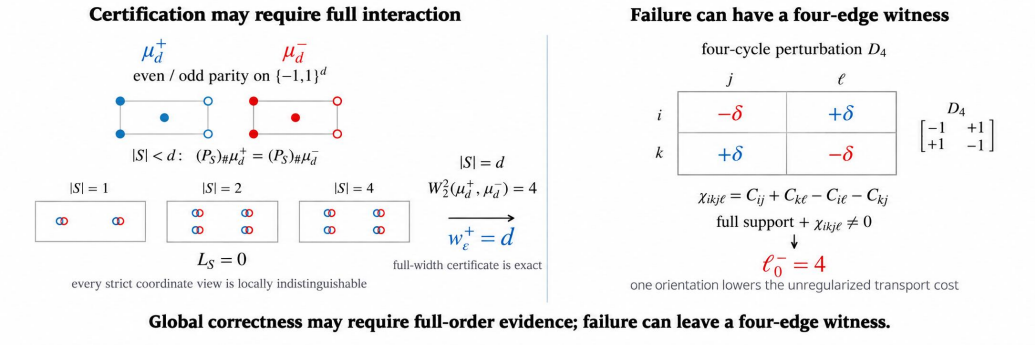


Figure 1: Verification asymmetry under a declared proof language. In the parity family, every strict coordinate view is locally indistinguishable, so coordinate-block certification requires full interaction width d. A full-support suboptimal plan, by contrast, is refuted by one improving 2×2 transport rectangle.

formalize *verification asymmetry*: positive evidence can require interaction order that grows with dimension while a valid counterexample remains constant-size. The parity family gives a sharp coordinate-block realization, $w_\varepsilon^+ = d$ versus $\ell_0^- = 4$, and the positive obstruction survives Wasserstein perturbations.

(2) Quantitative local-to-global guarantees. Safe local Kantorovich duals compose into global lower bounds, and orthogonal-block exactness is characterized by compatibility of local optima. Under explicit conditioning, a certificate gap controls distance to the optimal face. On the negative side, full-support optimality is rigid: some four-cycle has unit improvement margin at least half of the global suboptimality gap, with an exact squared-Euclidean geometric interpretation.

(3) A three-valued auditor and theorem-aligned evidence. VerifyOT returns CERTIFIED, REFUTED, or UNKNOWN under a common unregularized cost. A finite-query lower bound gives a concrete setting in which a sound verifier must abstain. Experiments calibrate the width hierarchy and four-cycle margin, test robustness and basis sensitivity, and audit actual Sinkhorn/Greenkhorn outputs without treating search failure as evidence of optimality.

2 PROBLEM SETUP AND VERIFICATION PROFILE

Let $x_1, \dots, x_n \in \mathbb{R}^d$ and $y_1, \dots, y_m \in \mathbb{R}^d$ have positive weights $a \in \Delta_n$ and $b \in \Delta_m$. For a cost matrix $C \in \mathbb{R}^{n \times m}$,

$$P^* := \min_{\gamma \in \Pi(a,b)} \langle C, \gamma \rangle, \qquad \Pi(a,b) := \{\gamma \geq 0 : \gamma \mathbf{1}_m = a,\ \gamma^\top \mathbf{1}_n = b\}. \tag{1}$$

A feasible candidate $\widehat{\gamma}$ has unregularized cost $U(\widehat{\gamma}) = \langle C, \widehat{\gamma} \rangle \geq P^*$. We call $\widehat{\gamma}$ *full-support* when $\widehat{\gamma}_{ij} > 0$ for every cell. A positive verifier constructs $L \leq P^*$; a negative verifier constructs a feasible zero-marginal perturbation D with $\widehat{\gamma} + D \geq 0$ and $\langle C, D \rangle < 0$.

Given tolerance $\varepsilon \geq 0$, VerifyOT returns CERTIFIED if $U - L \leq \varepsilon$, REFUTED if a replayed circulation decreases the cost by more than the required gain threshold, and UNKNOWN otherwise. A weak lower bound is not evidence against a candidate, and a failed local cycle search is not evidence for it.

Definition 1 (Positive verification width)**.** Fix an orthogonal basis and let $\mathfrak{G}_k$ be the family of orthogonal block partitions with maximum block dimension at most k. For a feasible candidate $\widehat{\gamma}$,

$$w_\varepsilon^+(\widehat{\gamma}; \mathfrak{G}) := \min\{k : \exists\, \mathcal{P} \in \mathfrak{G}_k,\ U(\widehat{\gamma}) - L_{\mathcal{P}} \leq \varepsilon\}, \tag{2}$$

with value $+\infty$ if the candidate is not ε-optimal.

Definition 2 (Negative refutation size)**.** For gain threshold $\tau \geq 0$, let $\ell_\tau^-(\widehat{\gamma})$ be the minimum number of edges in a capacity-feasible transport cycle whose replayed cost decrease exceeds τ, and $+\infty$ if no such cycle exists.

The pair $(w_\varepsilon^+, \ell_\tau^-)$ is the *bidirectional transport verification profile*. A proof language specifies the evidence a verifier may inspect, much as a query model specifies an algorithmic oracle. Positive width records the largest interaction order used by a certificate, whereas negative refutation size records the support of a replayable counterexample; we compare their scaling, not their numerical units. We say a family exhibits *unbounded verification asymmetry* when positive width grows with dimension while the refutation size of full-support suboptimal candidates remains bounded. The positive quantity is explicitly grammar-relative; the four-cycle results below are transport-structural once the candidate, cost, and full-support condition are fixed. The width is an oracle minimum over the grammar, while an implementation reports an *attained width* $\widehat{w}_\varepsilon^+ \geq w_\varepsilon^+$.

3 Positive Verification Complexity

3.1 The main separation

Let μ_d^+ and μ_d^- be the uniform distributions on even- and odd-parity vertices of $\{-1,1\}^d$:

$$\mu_d^\pm = \mathrm{Unif}\{x \in \{-1,1\}^d : \prod_{r=1}^d x_r = \pm 1\}. \tag{3}$$

Any even and odd vertex differ in at least one coordinate, while flipping one coordinate gives a cost-4 bijection under squared Euclidean cost. Hence $W_2^2(\mu_d^+, \mu_d^-) = 4$.

Theorem 3 (Sharp separation under coordinate-block proofs)**.** *Fix the coordinate-block grammar, $d \geq 3$, and $0 \leq \varepsilon < 4$. On the parity OT instance:*

1. *every optimal candidate has $w_\varepsilon^+ = d$;*
2. *every full-support suboptimal candidate has $\ell_0^- = 4$.*

Thus positive evidence must resolve the full d-way interaction, whereas negative evidence can remain constant-size.

Every strict coordinate marginal of μ_d^+ equals the corresponding marginal of μ_d^-. Hence every strict coordinate-block certificate has lower bound 0, although the global optimum is 4; the full block is exact. Section 4 supplies the other side: full support makes an improving four-cycle feasible, and four is the shortest nonzero circulation in a bipartite transport graph. Thus the positive evidence order grows with d, while the negative proof object keeps constant support. This is a scaling separation between two forms of admissible evidence, not a comparison of their numerical units.

3.2 Local duals as proof atoms

Consider measurable maps A_r, local costs c_r, and integrable dual potentials $\phi_r \circ A_r \in L^1(\mu)$, $\psi_r \circ A_r \in L^1(\nu)$ satisfying

$$\sum_{r=1}^R c_r(A_r x, A_r y) \leq c(x,y), \qquad \phi_r(A_r x) + \psi_r(A_r y) \leq c_r(A_r x, A_r y). \tag{4}$$

Theorem 4 (Compositional soundness)**.** *The sums $\Phi(x) = \sum_r \phi_r(A_r x)$ and $\Psi(y) = \sum_r \psi_r(A_r y)$ are feasible for the global Kantorovich dual. Their dual value is therefore a rigorous global lower bound.*

Local primal optima may disagree; only the verified dual inequalities need to compose. For quadratic OT, orthogonal blocks satisfy the cost-domination condition with equality.

Let $\mathcal{P} = \{E_1, \ldots, E_q\}$ be an orthogonal decomposition with projectors P_r, and define

$$L_{\mathcal{P}} := \sum_{r=1}^q W_2^2((P_r)_\# \mu, (P_r)_\# \nu). \tag{5}$$

For each block, let Γ_r^* be the full-space couplings whose projected coupling is block-optimal.

Proposition 5 (Compatibility criterion for exact block certificates)**.**

$$L_{\mathcal{P}} = W_2^2(\mu, \nu) \quad \Longleftrightarrow \quad \bigcap_{r=1}^{q} \Gamma_r^* \neq \varnothing. \tag{6}$$

For an optimal candidate, the exact block width is therefore the smallest maximum block dimension for which the local OT problems admit a common optimum.

This equivalence identifies compatibility, rather than projection accuracy alone, as the reason a low-width proof can be exact. Two blocks can each be solved perfectly yet still prefer incompatible matchings, in which case their summed lower bound remains strictly below the global OT value.

3.3 Positive hardness survives perturbations

The parity construction is exact, but the obstruction is not confined to a single pair of measures.

Theorem 6 (Robust positive hardness)**.** *Let $(\mu_0, \nu_0) = (\mu_d^+, \mu_d^-)$ and suppose $W_2(\mu, \mu_0) \leq \delta_\mu$ and $W_2(\nu, \nu_0) \leq \delta_\nu$, with $\delta = \delta_\mu + \delta_\nu$. For every coordinate partition $\mathcal{P}$ of width below d,*

$$L_{\mathcal{P}} \leq \delta^2, \qquad W_2^2(\mu, \nu) - L_{\mathcal{P}} \geq (2 - \delta)_+^2 - \delta^2. \tag{7}$$

In particular, for $0 \leq \delta < 1$ the certificate gap is at least $4(1 - \delta) > 0$.

Orthogonality makes the bound independent of the number of blocks: projected perturbation energies sum to at most the full Wasserstein perturbation energy. Low-order proofs therefore remain loose throughout a nontrivial Wasserstein neighborhood of the hard family. Appendix B stress-tests this statement over 120 perturbed instances.

3.4 From value evidence to plan evidence

Let $\mathcal{T} = \Pi(a, b)$ be the transportation polytope, $\mathcal{V}$ its vertices, and F^* its optimal face. Define

$$\Delta_{\text{vert}} := \min_{v \in \mathcal{V} \setminus F^*} (\langle C, v \rangle - P^*), \tag{8}$$

with $\Delta_{\text{vert}} = +\infty$ when all vertices are optimal.

Theorem 7 (Certificate gap controls plan error)**.** *For any feasible candidate $\widehat{\gamma}$ and valid lower bound L,*

$$\operatorname{dist}_1(\widehat{\gamma}, F^*) \leq \frac{2[U(\widehat{\gamma}) - L]}{\Delta_{\text{vert}}}. \tag{9}$$

If an optimal dual has reduced-cost margin $\rho > 0$ outside its zero-reduced-cost graph S^*, then

$$\sum_{(i,j) \notin S^*} \widehat{\gamma}_{ij} \leq \frac{U(\widehat{\gamma}) - L}{\rho}. \tag{10}$$

The conditioning terms are unavoidable: when different vertices are nearly tied, a small value gap cannot determine a unique correspondence. When the instance is separated, however, the certificate gap controls both distance to the optimal face and mass placed on definitely nonoptimal edges.

4 Local Refutation Complexity

The residual-cycle criterion is classical: a feasible transport plan is optimal iff its residual graph has no negative-cost directed cycle (Ahuja et al., 1993). Our question is narrower and quantitative: how short must a counterexample be, and how strongly can it witness the global error?

For $i \neq k$ and $j \neq \ell$, define the rectangular cross difference

$$\chi_{ikj\ell} := C_{ij} + C_{k\ell} - C_{i\ell} - C_{kj}. \tag{11}$$

Theorem 8 (Full-support rigidity and four-edge refutability)**.** *Assume $a_i, b_j > 0$ and $\gamma_{ij} > 0$ for every cell. The following are equivalent:*

1. *γ is globally optimal;*
2. *$C_{ij} = \alpha_i + \beta_j$ for some vectors α, β;*
3. *$\chi_{ikj\ell} = 0$ for every rectangle.*

Unless the cost is additively separable, every full-support coupling is therefore refutable by a four-cycle.

Full support is exactly what makes every oriented rectangle capacity-feasible: any nonzero cross difference has a cost-decreasing orientation and both removed edges carry positive mass. Without full support, residual-cycle optimality still applies, but missing support edges can force the shortest feasible improving cycle to be longer than four.

Corollary 9 (Euclidean geometry of refutation)**.** *If $C_{ij} = \|x_i - y_j\|^2$, then*

$$\chi_{ikj\ell} = 2\langle x_i - x_k,\, y_\ell - y_j\rangle. \tag{12}$$

A full-support coupling can be optimal only if $\operatorname{span}\{x_i - x_k\} \perp \operatorname{span}\{y_j - y_\ell\}$.

Thus exact full-support optimality under squared Euclidean cost is geometrically exceptional: a single nonorthogonal pair of source and target displacement directions yields a four-point contradiction.

4.1 A LOCAL MARGIN CONTROLLED BY THE GLOBAL GAP

Let $\widehat{\gamma}$ be full-support and let $\Delta := U(\widehat{\gamma}) - P^* > 0$.

Theorem 10 (Quantitative four-cycle theorem)**.** *There exists an oriented four-cycle D_Q with*

$$\langle C, D_Q\rangle \le -\frac{\Delta}{2}. \tag{13}$$

Let b_Q be the minimum mass on the two edges removed by that orientation. Then $\widehat{\gamma} + b_Q D_Q$ is feasible and

$$U(\widehat{\gamma}) - U(\widehat{\gamma} + b_Q D_Q) \ge \frac{b_Q \Delta}{2}. \tag{14}$$

Consequently the gain is at least $\eta\Delta/2$ for $\eta = \min_{ij} \widehat{\gamma}_{ij}$.

The quantitative consequence is local: some four-edge witness captures at least half of the global suboptimality per unit mass. The factor $1/2$ is a worst-case consequence of the ℓ_1 mass bound in the anchored rectangle decomposition; the empirical margins are much larger, and the optimal instance-dependent factor is open. The executable gain additionally depends on the removable mass b_Q, which can be much larger than the global minimum entry η.

The witness is stable under cost perturbations. Let $r(C) = \max_{i,k,j,\ell} |\chi_{ikj\ell}|$. If $\|\widetilde{C} - C\|_\infty \le \zeta$, then $r(\widetilde{C}) \ge r(C) - 4\zeta$. Hence a full-support plan remains four-cycle refutable whenever $r(C) > 4\zeta$.

5 INFORMATION LIMITS AND ABSTENTION

A verifier that observes only incomplete low-dimensional information cannot always know whether the missing directions matter.

Theorem 11 (Finite-query impossibility)**.** *Consider a deterministic verifier that adaptively queries pushforward OT problems through linear maps $A_1, \dots, A_T$. If their combined row span has rank below d, then there exist two OT instances with identical complete query transcripts but different global squared Wasserstein costs; one may have cost 0 and the other cost $\|v\|^2 > 0$.*

The construction hides a translation in the common kernel of all queried maps. Thus incomplete observations can make certification impossible for a sound deterministic verifier. VerifyOT therefore allows the third outcome UNKNOWN.

6 VERIFYOT

VerifyOT receives a candidate plan from any solver, repairs only numerical marginal error, evaluates the candidate under a common unregularized target cost, and runs positive and negative verification in parallel. The positive branch validates local dual feasibility and grows a block proof; the negative branch searches for four-cycles and, when needed, more general residual cycles. Every reported witness is replayed from the stored candidate and its cost decrease recomputed.

Input	Candidate coupling, target marginals, target cost, tolerance, and verification budget.
Positive branch	Compose safe local duals and increase attained width only when the current proof is insufficient.
Negative branch	Search replayable four-cycle or residual-cycle counterexamples.
Output	CERTIFIED with $[L, U]$, REFUTED with a circulation, or UNKNOWN.

Search versus intrinsic complexity. Theoretical width minimizes over the whole grammar; the implementation explores a finite search path. Greedy, PCA, seeded-random, and slack-guided proposals choose which atoms to test, while every accepted atom is independently checked before entering the proof.

Numerical safety. For a discrete dual (u, v), let $\xi = \max_{ij}[u_i + v_j - C_{ij}]_+$. Subtracting $\xi/2$ from each potential makes the pair feasible and reduces the unit-mass dual value by exactly ξ. A candidate with incorrect marginals is not used as an upper bound until it is repaired and rechecked.

Computational accounting. One-dimensional atoms use sorting and cumulative matching; small blocks use exact OT. Exhaustive four-cycle search is $O(n^2m^2)$, while randomized search is linear in its trial budget. Candidate generation, proof search, witness search, and replay are timed separately.

7 EXPERIMENTS

The experiments ask whether the theoretical evidence hierarchy appears numerically. We report the absolute certificate gap $G_k = U - L_k$; for exact candidates compared across datasets we also use the relative gap $G^\star_{\mathrm{rel}}(k) = (C^\star - L_k)/\max(C^\star, 10^{-12})$. All candidates are evaluated under the same unregularized squared-Euclidean cost, and experimental "width" means attained width unless theory identifies the oracle minimum. Failures, nonconvergence, timeouts, and UNKNOWN are retained.

Experimental setup. The registered campaign evaluates exact EMD, entropic Sinkhorn/Greenkhorn, sliced-transport candidates, and controlled corruptions; proof-critical checks use CPU float64 arithmetic. It ran on Ubuntu 22.04 with 16 Intel Xeon Platinum 8481C vCPUs, 80 GB RAM, and one RTX 4090D. Independent theorem calibration uses exact equal-weight CPU OT in a separate environment. We report these runs as verification evidence, not as end-to-end runtime benchmarks.

7.1 DIRECT CALIBRATION OF THE TWO CENTRAL THEOREM FAMILIES

Positive width behaves as predicted. Over product, parity, Gaussian, and Digits controls (10 seeds; widths $1, 2, 4, 8$), all 480 recorded lower bounds are valid and every nested hierarchy is monotone. For exact candidates, mean gaps are 0 at width one for product; $0.146, 0.146, 0.146, 0$ for parity; $0.295, 0.241, 0.141, 0$ for Gaussian; and $0.163, 0.139, 0.093, 0$ for Digits. Costs are computed after median-distance scaling, so the raw parity value 4 becomes $4/s^2$ while strict-width bounds remain zero. The invariant prediction is therefore the flat strict-width gap followed by closure at width d. At full width, the Sinkhorn gap matches $U - C^*$ to 3.1×10^{-16}.

The four-edge margin is not merely existential. We form full-support candidates

$$\widehat{\gamma}_\alpha = (1 - \alpha)\gamma^* + \alpha ab^\top, \qquad \alpha \in \{.01, .03, .1, .3, .5\}, \tag{15}$$

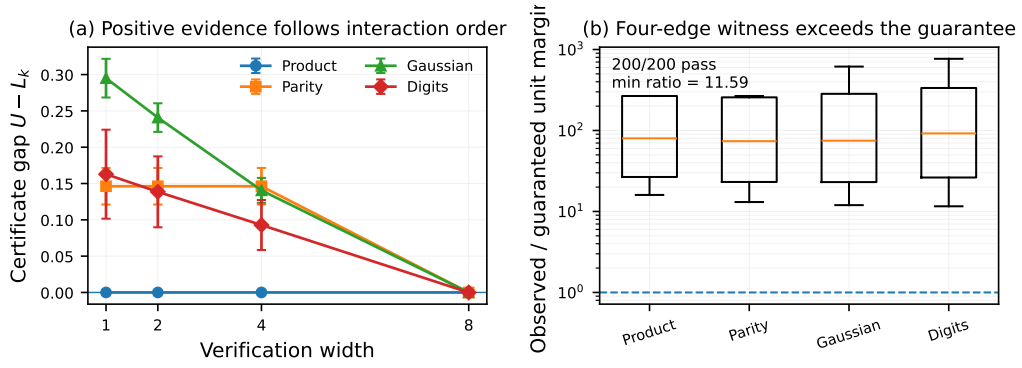


Figure 2: Direct theorem calibration over 10 seeds. **Left:** exact-candidate certificate gap versus block width for $n = 64, d = 8$. Product structure closes at width one; parity retains its entire gap through widths $1, 2, 4$ and closes only at full width; Gaussian and Digits tighten progressively. Error bars show one standard deviation. **Right:** ratio of the observed four-cycle unit margin to the theorem guarantee $\Delta/2$ for 200 full-support mixtures. All cases exceed the threshold 1; the smallest observed ratio is 11.59.

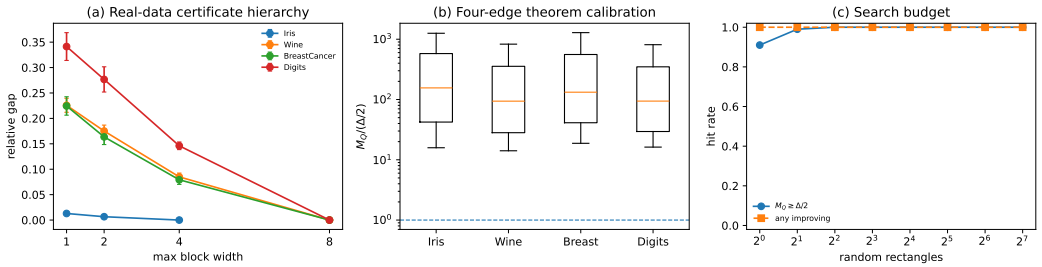


Figure 3: Small-data theorem validation (10 seeds). **Left:** exact block certificates tighten with width. **Middle:** all $200/200$ full-support mixtures satisfy the four-cycle margin bound. **Right:** random rectangles rapidly find theorem-strength witnesses on the same controlled mixtures.

for the same four data families and 10 seeds, giving 200 cases. Every case satisfies both $-\langle C, D_Q \rangle \geq \Delta/2$ and $\mathrm{gain}(Q) \geq b_Q \Delta/2$, with machine-precision marginal residuals. The minimum and median ratios to the theorem threshold are 11.59 and 80.0. Random rectangles reach the theorem-strength margin in 86.5%, 98.5%, and 100% of cases within one, two, and four draws. This is a controlled discoverability diagnostic, not a general search-time bound.

7.2 Registered Controls and Natural-Data Audit

The larger registered campaign is consistent with these controls: product cases close at width one in all $40/40$ formal units, while parity closure depends on the declared basis/search rule (planted, greedy, and joint-PCA: $30/30$ each; seeded-random: $10/30$). On planted corruptions, the negative branch refutes $2369/2640$ synthetic-core and $713/760$ preliminary Digits candidates at zero gain threshold. Full registered tables are deferred to Appendix B.

Real-data hierarchies. For exact class-pair transports, mean width-one relative gaps are 0.013 (Iris), 0.226 (Wine), 0.224 (Breast Cancer), and 0.341 (Digits). Every hierarchy is valid and monotone, and by width four each dataset has removed at least half of its width-one gap. On 200 full-support mixtures from the same datasets, all $200/200$ cases satisfy the quantitative four-cycle and executable-gain bounds; the minimum unit-margin ratio is 14.11. Appendix B reports the remaining robustness, basis, and search-budget diagnostics.

Actual Sinkhorn/Greenkhorn outputs. To remove the dependence on mixture construction, we audit log-domain Sinkhorn and Greenkhorn outputs on the same four datasets, with 10 seeds, 40

samples per side, and four regularization factors. The verifier performs only marginal-feasibility repair before recomputing all costs and witnesses. Table 1 separates mathematical evidence on the repaired candidate from a conservative 10^{-9} raw-solver convergence gate. At regularization .01, all $80/80$ repaired candidates satisfy the 1% certificate at full width; exhaustive coordinate-partition search finds no lower-width closure. Under the strict gate, 35 are counted CERTIFIED and 45 remain no-decision. At larger regularization no repaired candidate meets the 1% positive tolerance. All $320/320$ repaired candidates are full-support and admit an executable improving four-cycle at $\tau = 0$; all satisfy $M_Q \geq \Delta/2$, with minimum ratio 21.06. Random rectangles find a theorem-strength unit margin in 87.5%, 99.1%, and 100% of candidates within one, two, and four draws. No single cycle exceeds $0.01C^\star$, separating exact refutability from thresholded severity.

Table 1: Audit of actual Sinkhorn/Greenkhorn candidates, pooled over both solvers (80 cases per row). "C/U/ND" denotes CERTIFIED/UNKNOWN/no-decision under the strict raw-solver gate; repaired-candidate columns are evaluated after marginal feasibility repair.

$\lambda/\mathrm{med}(C_+)$	C/U/ND	Repair cert.	4-cycle, $\tau = 0$	$\overline{\Delta/C^\star}$
.01	$35/0/45$	$80/80$	$80/80$	.0047
.03	$0/71/9$	$0/80$	$80/80$	.0232
.10	$0/80/0$	$0/80$	$80/80$	.1048
.30	$0/80/0$	$0/80$	$80/80$	.2482

The broader registered audit uses a fixed LOCAL budget: across five datasets it yields $0/2200$ CERTIFIED, $647/2200$ REFUTED, $1436/2200$ completed UNKNOWN, and 117 separate no-decision rows. This does not contradict Table 1: the registered LOCAL budget restricts the proof search, whereas the follow-up audit measures how evidence changes when full coordinate-block width is allowed. Neither experiment identifies a basis-free intrinsic width or establishes a universal runtime advantage.

8 RELATED WORK

Approximate and structured transport. Entropic regularization and accelerated scaling make OT practical but change the returned coupling (Cuturi, 2013; Altschuler et al., 2017; Genevay et al., 2019). Minibatch approximations trade global consistency for local solves (Fatras et al., 2020; Nguyen et al., 2022). Low-rank transport imposes structural constraints on feasible couplings (Forrow et al., 2019; Scetbon et al., 2021; Scetbon & Cuturi, 2022), while neural OT learns maps or plans from data (Makkuva et al., 2020; Asadulaev et al., 2024). These methods motivate post-hoc auditing because the coupling itself is often consumed downstream.

Projected and observable transport. Sliced and projected OT replace or constrain a high-dimensional problem using low-dimensional transports (Paty & Cuturi, 2019; Muzellec & Cuturi, 2019; Lin et al., 2020; Mahey et al., 2023). Statistical analyses clarify when low-dimensional structure can mitigate high-dimensional Wasserstein estimation (Niles-Weed & Rigollet, 2022). Recent work lifts sliced transports to full plans (Liu et al., 2025; Tanguy et al., 2025), generalizes the slicing family (Chapel et al., 2025), or predicts full potentials from sliced potentials (Truong & Nguyen, 2026). Observable Wasserstein builds lower-bound hierarchies from restricted observables (Lopes dos Santos et al., 2026), and sliced-deficit rigidity studies when projection lower bounds are exact and stable (Han, 2026). Their primary objects are distances, generated plans, or equality structure; our width is candidate-specific and grammar-dependent.

Certificates and negative witnesses. Kantorovich duality already turns a feasible primal–dual pair into an optimality gap (Kantorovich, 1942; Villani, 2009; Peyré & Cuturi, 2019); computable certificates are classical in convex optimization (Nemirovski et al., 2010; Burns & Liang, 2026), and low-rank OT has rank-sufficiency conditions (Jawanpuria & Mishra, 2026). Residual-cycle optimality and transportation pivots are likewise classical (Ahuja et al., 1993). Our object is different: for a fixed candidate and declared evidence model, we ask for the interaction order of a positive proof and the support of a replayable counterexample. The sharp results use OT-specific structure–orthogonal Wasserstein decomposition, rectangular circulations, and the squared-Euclidean cross-

difference identity–to obtain the coordinate-block d-versus-four scaling separation and a four-cycle margin controlled by the global gap.

9 Discussion: What the Verification Profile Adds

What is grammar-relative, and what is transport-structural? The positive lower bound in Theorem 3 is deliberately not basis-free: it is a statement about the coordinate-block language, just as a query lower bound is stated for a specified oracle. Appendix B shows empirically that changing the basis changes attained certificates. By contrast, Theorems 8 and 10 do not depend on the positive proof grammar. Once the cost and a full-support candidate are fixed, four-edge refutability and its margin follow from the geometry of the transportation polytope. The main separation therefore combines a grammar-relative positive lower bound with a transport-structural negative upper bound under an explicit support condition.

Evidence size is not search cost. The profile minimizes over admissible evidence, whereas VerifyOT explores only a finite search path. The real-solver audit makes this distinction visible: at regularization .01, exhaustive coordinate-partition search closes the 1% gap only at full width, while the registered LOCAL budget attains no positive closure. Thus $0/2200$ CERTIFIED measures that budget, not an intrinsic impossibility. Conversely, the random-rectangle hit rates show that short witnesses are easy to find in the tested full-support candidates, but they are not worst-case search-complexity guarantees.

How to read the full-support assumption. Full support is strong, but it matches the entropic candidates audited in Table 1 after feasibility repair. It is not imposed on sparse OT plans. For a sparse candidate, residual-cycle optimality remains exact, yet missing support edges can make the shortest feasible improving cycle longer than four. A support-sensitive characterization of refutation length is therefore a natural extension of the present profile.

Why the sharp results are transport-specific. A bidirectional evidence profile could be defined for a generic linear program, but the present separation does not follow from LP duality alone. Orthogonal Wasserstein costs provide additive positive proof atoms, transportation polytopes provide rectangular zero-marginal circulations, and squared-Euclidean costs turn each rectangle score into a displacement inner product. Together these structures yield the full-order positive obstruction and the $\Delta/2$ four-cycle margin in the same problem class.

Operational contract. VerifyOT is an auditor rather than a replacement solver. A CERTIFIED output carries a checked numerical gap, a REFUTED output carries a replayable feasible cost decrease, and UNKNOWN makes no optimality claim. This contract remains meaningful when a low-width proof does not close.

10 Limitations and Conclusion

Limitations. Verification width and refutation size measure evidence rather than search cost. Width depends on proof language, and the finite-query lower bound covers deterministic verifiers. The four-edge guarantee requires full support; sparse candidates satisfy residual-cycle optimality but may need longer witnesses. Guarantees depend on conditioning such as Δ_{vert}, while refutation concerns exact optimality for unregularized cost, so an entropic plan may remain useful but refutable. Our experiments do not establish universal low-width coverage or runtime gains.

Conclusion. OT verification exhibits a proof-language-relative asymmetry: positive evidence can require full interaction order, while a full-support suboptimal plan admits a four-edge witness whose unit margin reflects global error. VerifyOT makes this distinction operational through a checked certificate, a replayable counterexample, or UNKNOWN. The real-solver audit shows both sides on solver-produced candidates: positive closure near the exact-OT regime and constant-size exact refutations after feasibility repair. The broader objective is not to replace OT solvers, but to make the evidence supporting or contradicting a returned coupling explicit.

AI USE STATEMENT

Generative AI tools were used solely for language polishing and formatting assistance. All research ideas, methods, mathematical statements, proofs, citations, experimental protocols, code, and numerical claims were developed, conducted, and verified by the authors. The authors take full responsibility for the final content of this work.

ETHICS STATEMENT

The work studies numerical verification of optimal-transport couplings and does not make clinical, legal, or high-stakes decisions. Public medical and single-cell datasets are used only as numerical benchmark domains. Dataset licenses and privacy requirements must be followed. A VerifyOT certificate concerns the stated transport objective; it does not by itself certify fairness, causal validity, or downstream safety.

REPRODUCIBILITY STATEMENT

Complete proofs are provided in Appendix A. Appendix B reports the experimental protocols and additional analyses for robustness, basis sensitivity, registered status composition, and the Sinkhorn/Greenkhorn audit. It specifies the preprocessing and solver settings used in the actual-solver audit and preserves the distinction between verification decisions and run status, including UNKNOWN, timeout, nonconvergence, repair failure, and candidate errors. Oracle-assisted `FULL_REFERENCE` results are reported separately from LOCAL results.

REFERENCES

Ravindra K. Ahuja, Thomas L. Magnanti, and James B. Orlin. *Network Flows: Theory, Algorithms, and Applications*. Prentice Hall, 1993.

Jason Altschuler, Jonathan Weed, and Philippe Rigollet. Near-linear time approximation algorithms for optimal transport via sinkhorn iteration. In *Advances in Neural Information Processing Systems*, volume 30, 2017.

Arip Asadulaev, Aleksandr Korotin, Vage Egiazarian, Petr Mokrov, and Evgeny Burnaev. Neural optimal transport with general cost functionals. In *International Conference on Learning Representations*, 2024.

Matthew X. Burns and Jiaming Liang. Accuracy certificates for convex optimization at accelerated rates via primal-dual averaging, 2026.

Laetitia Chapel, Romain Tavenard, and Samuel Vaiter. Differentiable generalized sliced wasserstein plans. In *Advances in Neural Information Processing Systems*, volume 38, 2025.

Marco Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. In *Advances in Neural Information Processing Systems*, 2013.

Kilian Fatras, Younes Zine, Rémi Flamary, Rémi Gribonval, and Nicolas Courty. Learning with minibatch wasserstein: Asymptotic and gradient properties. In *Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics*, pp. 2131–2141, 2020.

Aden Forrow, Jan-Christian Hütter, Mor Nitzan, Philippe Rigollet, Geoffrey Schiebinger, and Jonathan Weed. Statistical optimal transport via factored couplings. In *Proceedings of the Twenty-Second International Conference on Artificial Intelligence and Statistics*, pp. 2454–2465, 2019.

Aude Genevay, Lénaïc Chizat, Francis Bach, Marco Cuturi, and Gabriel Peyré. Sample complexity of sinkhorn divergences. In *Proceedings of the Twenty-Second International Conference on Artificial Intelligence and Statistics*, pp. 1574–1583, 2019.

Bang-Xian Han. Rigidity and quantitative stability of the sliced wasserstein deficit, 2026.

Pratik Jawanpuria and Bamdev Mishra. A riemannian approach to low-rank optimal transport. In *Machine Learning and Knowledge Discovery in Databases: ECML PKDD 2026*, pp. 579–597, 2026.

Leonid V. Kantorovich. On the translocation of masses. *Doklady Akademii Nauk USSR*, 37:199–201, 1942.

Tianyi Lin, Chenyou Fan, Nhat Ho, Marco Cuturi, and Michael I. Jordan. Projection robust wasserstein distance and riemannian optimization. In *Advances in Neural Information Processing Systems*, volume 33, 2020.

Xinran Liu, Rocio Diaz Martin, Yikun Bai, Ashkan Shahbazi, Matthew Thorpe, Akram Aldroubi, and Soheil Kolouri. Expected sliced transport plans. In *International Conference on Learning Representations*, 2025.

Edivaldo Lopes dos Santos, Leandro Vicente Mauri, Washington Mio, and Tom Needham. The observable wasserstein distance, 2026.

Guillaume Mahey, Laetitia Chapel, Gilles Gasso, Clément Bonet, and Nicolas Courty. Fast optimal transport through sliced generalized wasserstein geodesics. In *Advances in Neural Information Processing Systems*, volume 36, 2023.

Ashok Vardhan Makkuva, Amirhossein Taghvaei, Sewoong Oh, and Jason D. Lee. Optimal transport mapping via input convex neural networks. In *Proceedings of the 37th International Conference on Machine Learning*, pp. 6672–6681, 2020.

Boris Muzellec and Marco Cuturi. Subspace detours: Building transport plans that are optimal on subspace projections. In *Advances in Neural Information Processing Systems*, volume 32, 2019.

Arkadi Nemirovski, Shmuel Onn, and Uriel G. Rothblum. Accuracy certificates for computational problems with convex structure. *Mathematics of Operations Research*, 35(1):52–78, 2010.

Khai Nguyen, Dang Nguyen, Quoc Dinh Nguyen, Tung Pham, Hung Bui, Dinh Phung, Trung Le, and Nhat Ho. On transportation of mini-batches: A hierarchical approach. In *Proceedings of the 39th International Conference on Machine Learning*, pp. 16622–16655, 2022.

Jonathan Niles-Weed and Philippe Rigollet. Estimation of wasserstein distances in the spiked transport model. *Bernoulli*, 28(4):2663–2688, 2022.

François-Pierre Paty and Marco Cuturi. Subspace robust wasserstein distances. In *Proceedings of the 36th International Conference on Machine Learning*, pp. 5072–5081, 2019.

Gabriel Peyré and Marco Cuturi. Computational optimal transport. *Foundations and Trends in Machine Learning*, 11(5–6):355–607, 2019.

Meyer Scetbon and Marco Cuturi. Low-rank optimal transport: Approximation, statistics and debiasing. In *Advances in Neural Information Processing Systems*, volume 35, 2022.

Meyer Scetbon, Marco Cuturi, and Gabriel Peyré. Low-rank sinkhorn factorization. In *Proceedings of the 38th International Conference on Machine Learning*, pp. 9344–9354, 2021.

Eloi Tanguy, Laetitia Chapel, and Julie Delon. Sliced optimal transport plans, 2025.

Minh-Phuc Truong and Khai Nguyen. Amortized optimal transport from sliced potentials, 2026.

Cédric Villani. *Optimal Transport: Old and New*. Springer, 2009.

A Proofs and additional derivations

A.1 Proof of Theorem 4

For every x, y, local feasibility and equation 4 give

$$\Phi(x) + \Psi(y) = \sum_r \big[\phi_r(A_r x) + \psi_r(A_r y)\big] \le \sum_r c_r(A_r x, A_r y) \le c(x, y).$$

Thus (Φ, Ψ) is globally dual-feasible. Integrating against the two marginals gives a dual lower bound by Kantorovich weak duality. No compatibility assumption on the local primal optimizers is used.

A.2 Proof of Theorem 5

For any coupling γ and orthogonal decomposition, quadratic cost splits as

$$C(\gamma) = \sum_r C_r(\gamma), \qquad C_r(\gamma) := \int \|P_r(x - y)\|^2 \, d\gamma.$$

Hence

$$P^* = \min_\gamma \sum_r C_r(\gamma) \ge \sum_r \min_\gamma C_r(\gamma) = L_{\mathcal{P}}.$$

If a coupling belongs to every Γ_r^*, it attains every local minimum and therefore attains the sum $L_{\mathcal{P}}$, proving equality. Conversely, if equality holds and γ^* is globally optimal, then each difference $C_r(\gamma^*) - \min_\gamma C_r(\gamma)$ is nonnegative and their sum is zero; every difference is therefore zero, so $\gamma^* \in \cap_r \Gamma_r^*$.

A.3 Proof of Theorem 3

For any strict coordinate subset $S \subsetneq [d]$, both parity distributions project to the uniform distribution on $\{-1, 1\}^{|S|}$. Every block cost in a partition of width $< d$ is therefore zero, so $L_{\mathcal{P}} = 0$. The global cost is 4: every even and odd vertex differ in at least one coordinate, giving cost at least 4, and flipping one coordinate gives a bijection with cost exactly 4. Thus no width-$< d$ block certificate can certify an optimal candidate to tolerance $\varepsilon < 4$, while the full block does, so $w_\varepsilon^+ = d$.

The second part follows from Theorem 8 (or the quantitative Theorem 10): every full-support sub-optimal coupling has an improving four-cycle. Any nonzero circulation on a bipartite graph has cycle length at least four, hence $\ell_0^- = 4$.

A.4 Proof of Theorem 6

The triangle inequality and $W_2(\mu_0, \nu_0) = 2$ give

$$W_2(\mu, \nu) \ge (2 - \delta)_+.$$

For a strict coordinate block S_r, the parity pushforwards coincide. Define

$$a_r := W_2((P_r)_\# \mu, (P_r)_\# \mu_0), \qquad b_r := W_2((P_r)_\# \nu, (P_r)_\# \nu_0).$$

The triangle inequality gives $W_2((P_r)_\# \mu, (P_r)_\# \nu) \le a_r + b_r$. For any coupling of μ and μ_0, orthogonality implies that the sum of projected squared displacements equals the full squared displacement. Taking an optimal full-space coupling therefore yields

$$\sum_r a_r^2 \le W_2^2(\mu, \mu_0) \le \delta_\mu^2, \qquad \sum_r b_r^2 \le \delta_\nu^2.$$

Hence, by Minkowski in ℓ_2,

$$\sqrt{L_{\mathcal{P}}} \le \Big(\sum_r (a_r + b_r)^2 \Big)^{1/2} \le \delta_\mu + \delta_\nu = \delta,$$

so $L_{\mathcal{P}} \le \delta^2$. Combining the two bounds proves equation 7. For $0 \le \delta < 1$, the right-hand side equals $4 - 4\delta > 0$.

A.5 Proof of Theorem 7

Write a feasible candidate as a convex combination of transportation-polytope vertices, $\widehat{\gamma} = \sum_{v \in \mathcal{V}} \lambda_v v$. Let α be the total weight on nonoptimal vertices. The primal gap is at least $\alpha \Delta_{\mathrm{vert}}$. The optimal-weight part can be normalized to a point of F^* when $\alpha < 1$; when $\alpha = 1$, choose any point of F^*. Because any two probability couplings have ℓ_1 distance at most 2,

$$\mathrm{dist}_1(\widehat{\gamma}, F^*) \le 2\alpha \le \frac{2[U(\widehat{\gamma}) - P^*]}{\Delta_{\mathrm{vert}}} \le \frac{2[U(\widehat{\gamma}) - L]}{\Delta_{\mathrm{vert}}}.$$

For the reduced-cost statement, let (u^*, v^*) be an optimal dual and $r_{ij} = C_{ij} - u_i^* - v_j^* \ge 0$. Then

$$U(\widehat{\gamma}) - P^* = \sum_{ij} r_{ij} \widehat{\gamma}_{ij} \ge \rho \sum_{(i,j) \notin S^*} \widehat{\gamma}_{ij},$$

and replace the true gap by the larger certificate gap.

A.6 Proof of Theorem 8 and Corollary 9

If a full-support coupling γ is optimal, complementary slackness with an optimal dual (u, v) gives $C_{ij} = u_i + v_j$ for every cell because every $\gamma_{ij} > 0$. This is additive separability. Additive separability makes all rectangular cross differences zero. Conversely, vanishing cross differences imply additive separability by fixing one reference row and column; then every feasible coupling has the same objective value $a^\top \alpha + b^\top \beta$, so every coupling is optimal. If any cross difference is nonzero, one orientation of the associated four-cycle has negative unit cost, and full support makes a positive step feasible.

For $C_{ij} = \|x_i - y_j\|^2$, direct expansion yields

$$C_{ij} + C_{k\ell} - C_{i\ell} - C_{kj} = 2\langle x_i - x_k, y_\ell - y_j \rangle.$$

All cross differences vanish exactly when the two difference spans are orthogonal.

A.7 Proof of Theorem 10

Let $H = \gamma^* - \widehat{\gamma}$, where γ^* is optimal. Then $H\mathbf{1} = 0$, $H^\top \mathbf{1} = 0$, $\|H\|_1 \le 2$, and $\langle C, H \rangle = -\Delta$. Fix the last row and column as anchors. For $i < n$ and $j < m$, let D^{ij} have $+1$ at (i, j) and (n, m), -1 at (i, m) and (n, j), and zero elsewhere. The zero-marginal constraints imply

$$H = \sum_{i<n, j<m} H_{ij} D^{ij}, \qquad \sum_{i<n, j<m} |H_{ij}| \le \|H\|_1 \le 2.$$

Orient each nonzero rectangle by $\mathrm{sgn}(H_{ij})$. If every oriented rectangle had unit cost greater than $-\Delta/2$, then

$$\langle C, H \rangle = \sum_{i<n, j<m} |H_{ij}| \, \langle C, \mathrm{sgn}(H_{ij}) D^{ij} \rangle > -\frac{\Delta}{2} \sum_{i<n, j<m} |H_{ij}| \ge -\Delta,$$

a contradiction. Thus some oriented four-cycle D_Q satisfies equation 13. Let b_Q be the smaller mass on its two removed edges. Full support gives $b_Q > 0$, so $\widehat{\gamma} + b_Q D_Q$ is feasible and its cost drops by at least $b_Q \Delta/2$. Since $b_Q \ge \eta := \min_{ij} \widehat{\gamma}_{ij}$, the uniform corollary follows.

A.8 Proof of Theorem 11

Run the deterministic adaptive verifier on the zero instance (δ_0, δ_0). Its answers determine the query sequence $A_1, \ldots, A_T$. Because their combined row span has rank below d, choose $v \ne 0$ in the common kernel. On (δ_0, δ_v) every queried pushforward is identical to that of the zero instance, so the verifier follows the same adaptive transcript. The true squared Wasserstein costs are 0 and $\|v\|^2$, respectively. Any decision depending only on the transcript must therefore be invalid for at least one instance if it claims a uniformly smaller uncertainty interval.

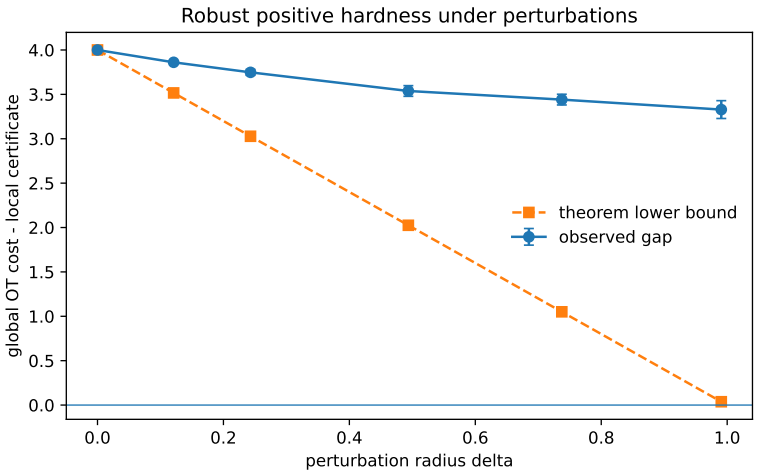


Figure 4: Robust positive hardness beyond the exact parity construction. Observed low-width gaps remain above the perturbation lower bound in every run.

B ADDITIONAL EXPERIMENTAL ANALYSES

Robust positive hardness. We perturb the even/odd parity supports with isotropic Gaussian noise over six perturbation levels and 10 seeds. Across 120 width-one/two records, every observed certificate gap remains above the lower bound in Theorem 6. At the largest tested average perturbation radius (approximately $\delta = 0.99$), the empirical gap remains far from zero, while the worst-case theorem bound is deliberately conservative near the edge of its nontrivial regime. Figure 4 shows more than numerical soundness: the empirical gap decays much more slowly than the worst-case guarantee. At the largest tested average radius, $\delta \approx 0.99$, the observed mean gap is about 3.33 while the theorem lower bound is about 0.04. The robust theorem is therefore conservative near the edge of its nontrivial regime, but the hard positive-verification phenomenon remains pronounced.

Basis and grammar sensitivity. On Iris, Wine, Breast Cancer, and Digits, we compare width-one certificates in the joint-PCA basis with 20 random orthogonal/subspace bases per seed. PCA yields a smaller mean gap than the random-basis mean on all four datasets, but random bases occasionally do better (most often on Breast Cancer). The result supports the paper's grammar-dependent interpretation of *attained* width rather than a claim that any particular basis is universally optimal. Figure 5 makes the grammar dependence operational. The PCA basis lowers the mean width-one gap relative to the random-basis mean on all four datasets, yet random bases outperform PCA in some draws (about 13%, 32%, 56%, and 11% for Iris, Wine, Breast Cancer, and Digits). Verification width should therefore be read relative to the declared proof language, not as a basis-free property of the data.

Actual-solver audit. The follow-up audit uses Iris, Wine, Breast Cancer, and Digits, 10 seeds, and 40 samples per side. Source and target samples are jointly standardized, expressed in a joint PCA proof basis ($d = 4$ for Iris and $d = 8$ otherwise), and median-distance scaled. Log-domain Sinkhorn and Greenkhorn are run at regularization factors $.01, .03, .10, .30$ times the median positive squared cost. Positive certification uses $\varepsilon = .01C^\star$ and coordinate-block widths $1, 2, 4, \ldots, d$; the negative audit exhaustively replays the best feasible four-cycle. A marginal-only IPF repair is applied before a candidate is used as a feasible upper bound.

All 320 repaired candidates have full support in float64. The largest ℓ_1 change introduced by marginal repair is 6.74×10^{-4}, and the largest raw marginal ℓ_∞ error is 7.79×10^{-6}. Every candidate admits an executable improving four-cycle, and every measured unit margin satisfies Theorem 10; the minimum and median ratios $M_Q/(\Delta/2)$ are 21.06 and 179.93. For the $.01$ runs, all 80 repaired candidates certify at full width, and exhaustive partition search confirms that none closes at a smaller maximum block width. Random rectangle search attains the theorem-strength margin

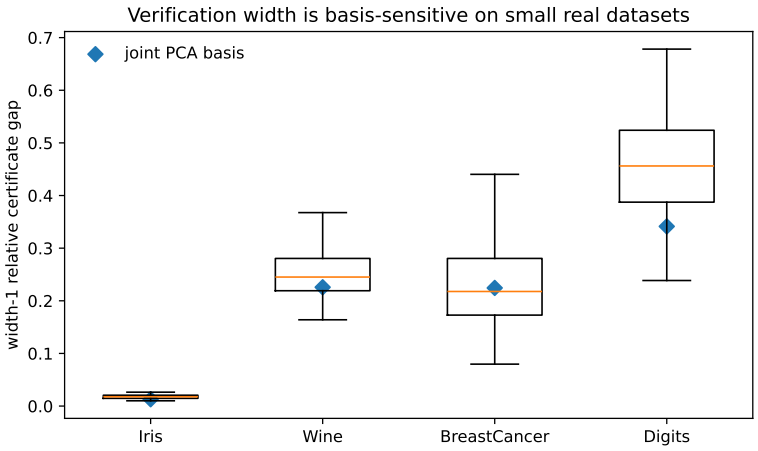


Figure 5: Basis sensitivity of width-one certificates on four small real datasets. Lower is stronger. The marked PCA basis improves the mean gap but is not uniformly optimal.

Table 2: Solver-specific results for the actual-output audit. Each row contains 40 candidates. Status uses the strict 10^{-9} raw-solver convergence gate; "exact ref." is four-cycle refutation at $\tau = 0$.

Solver	λ	Conv.	Cert.	Unk.	No-dec.	Exact ref.
Greenkhorn	.01	16	16	0	24	40/40
Greenkhorn	.03	35	0	35	5	40/40
Greenkhorn	.10	40	0	40	0	40/40
Greenkhorn	.30	40	0	40	0	40/40
Sinkhorn-log	.01	19	19	0	21	40/40
Sinkhorn-log	.03	36	0	36	4	40/40
Sinkhorn-log	.10	40	0	40	0	40/40
Sinkhorn-log	.30	40	0	40	0	40/40

in 87.5%, 99.1%, and 100% of candidates within one, two, and four queries, respectively, but no single rectangle achieves replay gain above $.01C^\star$.

Registered status composition. The five-dataset registered campaign retains all completed UNKNOWN decisions and separates them from execution/candidate-stage no-decision rows. FULL-REFERENCE uses exact-OT information and is shown only as an oracle-assisted diagnostic. Solver severity in the oracle-assisted diagnostic moves in the expected direction: increasing regularization from .01 to .1 gives paired certification/refutation effects of $0.800/0.810$ for Greenkhorn and $0.815/0.815$ for Sinkhorn-log. These are fixed-seed-block bootstrap effects under `FULL_REFERENCE`, not deployable coverage or runtime claims.

Random-rectangle search-budget stress tests. On the original 200 A2 full-support mixtures, the theorem-strength hit rate was 86.5% with one rectangle, 98.5% with two, and 100% by four. On the four additional real small-data families in Fig. 3, the corresponding rates are 91%, 99%, and 100%. These controlled-mixture results are not used as general search-complexity claims. On the 320 actual solver candidates, the full random-query curve is:

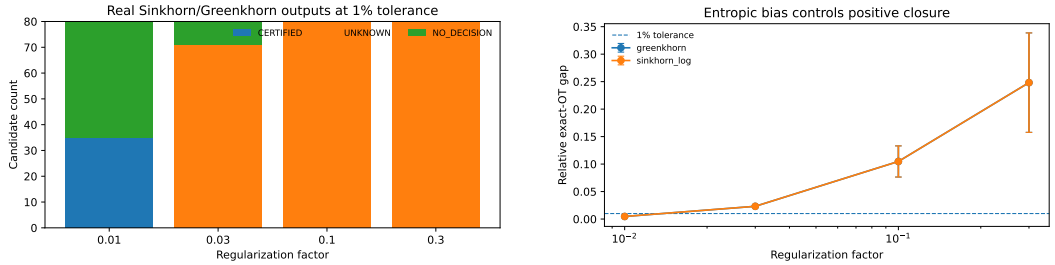


Figure 6: Actual-solver audit. **Left:** strict-gate decision composition by regularization. **Right:** solver suboptimality under the common unregularized cost. The figures separate raw solver convergence from mathematical checks on the repaired feasible candidate.

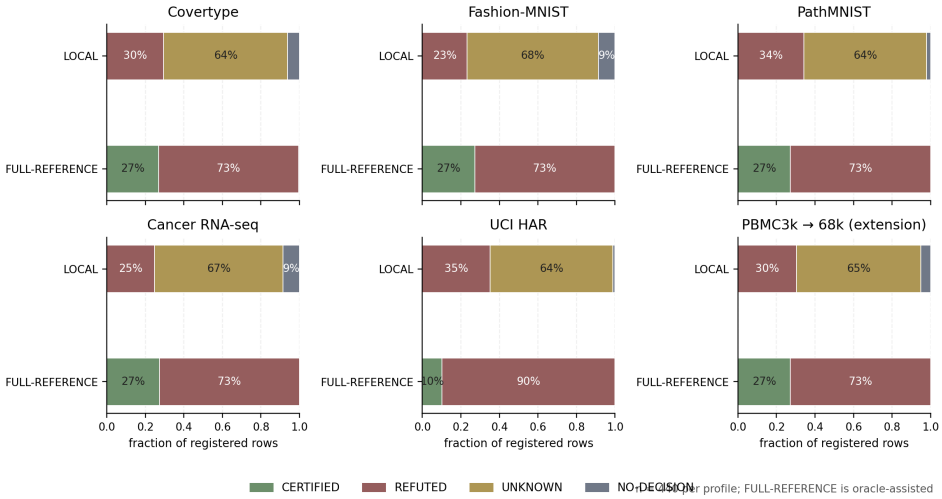


Figure 7: Registered status composition. LOCAL uses the fixed local budget; FULL-REFERENCE is oracle-assisted. PBMC is an extension and is not pooled into the five-dataset confirmatory macro.

Queries	$M_Q \geq \Delta/2$	Any gain	Gain $> .01C^\star$
1	87.5%	75.9%	0%
2	99.1%	82.8%	0%
4	100%	87.2%	0%
8	100%	93.4%	0%
16	100%	97.8%	0%
32	100%	99.4%	0%
64	100%	100%	0%
128	100%	100%	0%

The theorem-strength column concerns unit margin rather than executable gain; the last column shows why exact refutability and a 1% severity threshold should not be conflated.

The formal evidence preserves the distinction between verification decisions and run status. Candidate-stage errors, nonconvergence markers, residual-search timeouts, repair failures, and no-decision rows are not pooled into a single failure probability. The five-dataset confirmatory macro excludes PBMC, which is treated as an independent extension. `FULL_REFERENCE` uses exact-OT information and is never interpreted as deployable coverage or a runtime result. Development localization panels are diagnostic and are not promoted to formal ten-seed localization claims.